Qualifying Quality - A Framework for Supporting Quality-Informed Decisions

Discussion Paper

Bill Allen

Australian Bureau of Statistics

June 2002

Table of Contents

"We assist and encourage informed decision-making, research and discussion within governments and the community, by providing a high quality, objective and responsive national statistical service."

Mission Statement for the Australian Bureau of Statistcs

1. Introduction

The role of the Australian Bureau of Statistics (ABS) is often interpreted in the context of disseminating important social and economic data, which then form the basis for informed decisions. These decisions can then be considered 'data-informed decisions'. While this is a worthy and admirable role for the ABS, this paper proposes an alternative paradigm.

This new paradigm parallels a change in the interpretation in the concept of quality towards a broader concept of quality, where quality is defined as 'fitness for purpose' and is judged by the user. In keeping with this change in focus, the role of the ABS needs to consider how to assist the user in both making this assessment and applying the results of that assessment appropriately. This is not to suggest that the ABS needs to sit hand-in-hand with the user, verifying the way the data are being used is appropriate, but rather considers how the ABS can assist the user through providing the user with the necessary information and education for them in how to use that information. In doing so, the user is prompted to consider not only the data values, but also the quality of the data. Thus, the decisions evolve from 'data-informed decisions' to 'quality-informed decisions'.

This paper describes a 'decision cycle' which is used to provide a framework for the process of making a quality-informed decision. Each stage of the decision cycle is overviewed, followed by a discussion of the corresponding role for the ABS. Two primary tools are also introduced: the Quality Declaration; and the Quality Assessment. The Quality Declaration is used to document the quality of a data source, while the Quality Assessment is use to document the fitness for purpose of a data source against a specific data need. Together, they provide the user with the basic information necessary to make a quality-informed decision through the application of appropriate risk management strategies.

2. Executive Summary

The proposed framework looks at the role of data and the quality of data from the perspective of the data user and their underlying decision-making processes. In doing so, it highlights the importance of properly defining data needs, making available descriptions of the quality of data through Quality Declarations, comparing the identified data need with the data source as part of a Quality Assessment and implementing appropriate risk management strategies into the decision-making process to take into account that the data need and data sources do not perfectly align.

The role of the ABS is discussed in the light of this framework, both in identifying existing activity and in considering opportunities to focus and build upon this activity to better facilitate quality-informed decisions in government and the community. This can be broadly summarised as follows:

- In terms of defining data needs, the ABS can assist through both general user education (e.g. seminars and training courses) and targeted direct ABS involvement through avenues such as ABS outposted officers, subject areas and user consultation.

- In describing existing data sources, the ABS has an important role to play both in completing and disseminating Quality Declarations for ABS data and through disseminating guidelines and templates so that other data providers are well placed to document their own data collections using Quality Declarations.

- Finally, the ABS also has a key role in the risk management process through the quality assurance of ABS collections and in assisting potential users of the data is the appropriate use of data in the decision-making process.

3. The Decision Cycle

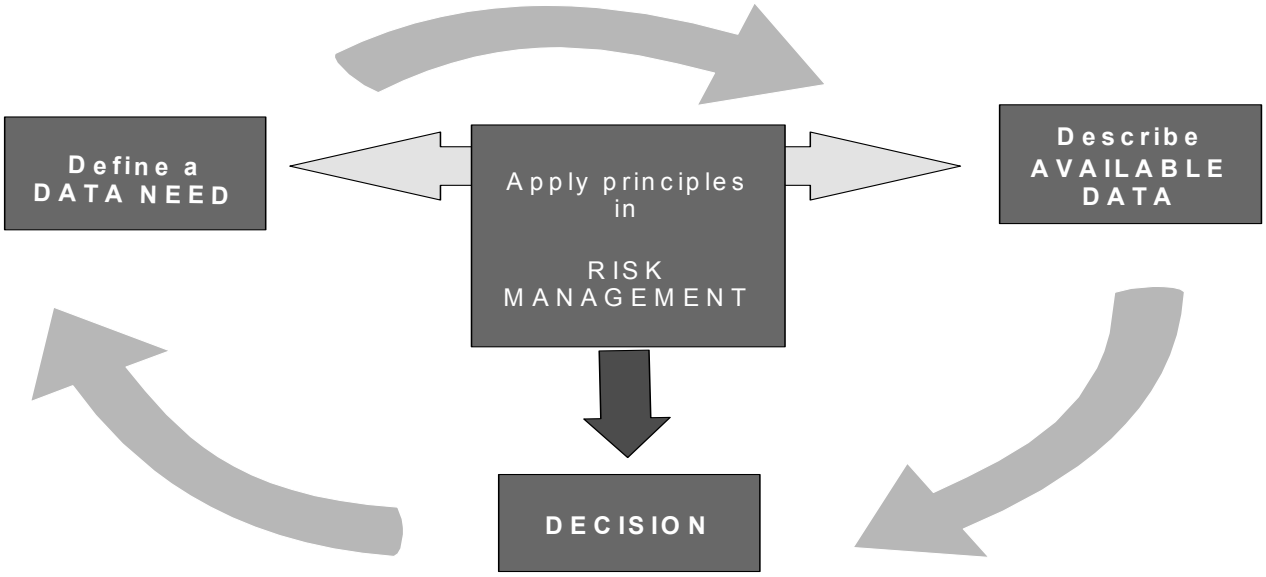
The decision cycle refers to the process which starts with a person wanting to make some underlying decision and finishes with the decision being made. A typical but non-preferred version of the decision cycle might go something along the lines of:

- Need to make a decision
- Want the decision to be based on data
- Look for data until something that looks close enough is found
- Use the data (as if it were perfect)

This decision cycle is focused on assisting in data-informed decision-making, but fails to take account of the quality of the data. Only a crude assessment is made of the fitness for purpose of the data and the degree to which the data are fit for purpose plays no role in the decision-making process.

Figure 1, below, describes a decision cycle proposed to facilitate quality-informed decision-making.

Figure 1 - The Decision Cycle



The decision cycle starts with the decision. The decision defines the desired outcome and the specific data need. This data need should be clearly defined in terms of the information required to facilitate the underlying decision. In essence, it describes the perfect data source for the decision required.

The next stage is to describe existing data sources. This description is called a Quality Declaration and forms the basis for the next stage as quality-informed decisions require an understanding of the quality (i.e. fitness for purpose) of the existing data sources. For the decision-maker, ideally, the range of possible existing data sources have already been appropriately documented and all they need do is reference the Quality Declarations for use in the next stage.

The final stage centres around the application of the data to the decision. The data need and data sources are compared against each other and mismatches are identified as the fitness of the data sources is assessed against the specific requirements of the underlying decision. This information is recorded in a Quality Assessment. Greater mismatches infer greater risks in using the data, so risk management principles are adopted which influence the final decision. Thus, just as the values of the data are used to make the final decision, the information about the quality of the data are also incorporated into the decision-making process to produce a quality-informed decision.

4. Defining Data Needs

The process of defining data needs draws its strength from two underlying models. The first model is the Inputs-Transformations-Outcomes (ITO) Model and is used to focus the data need on the underlying decision. The second model is a data quality framework and provides structure to the process of describing a data need.

The process of first defining the data need independent of existing data sources and

then comparing the need against existing data sources is essential to avoid the common mistake of defining data needs according to what data sources are available. Thus any shortcomings in the data are explicitly identified, considered and acted upon.

The process of defining a data need is detailed further under **4.3 Framework for defining data needs**.

4.1 The Inputs-Transformations-Outcomes (ITO) Model

The ITO model was developed by John Smyrk of the Australian National University and has been incorporated into ABS's project management framework. The model defines a project structure in terms of its objectives (outcomes), its deliverables (outputs) and how the project inputs are transformed, via outputs, into outcomes. Outcomes remain the ultimate objective, whereas the outputs are the physical deliverables that help achieve the outcomes.

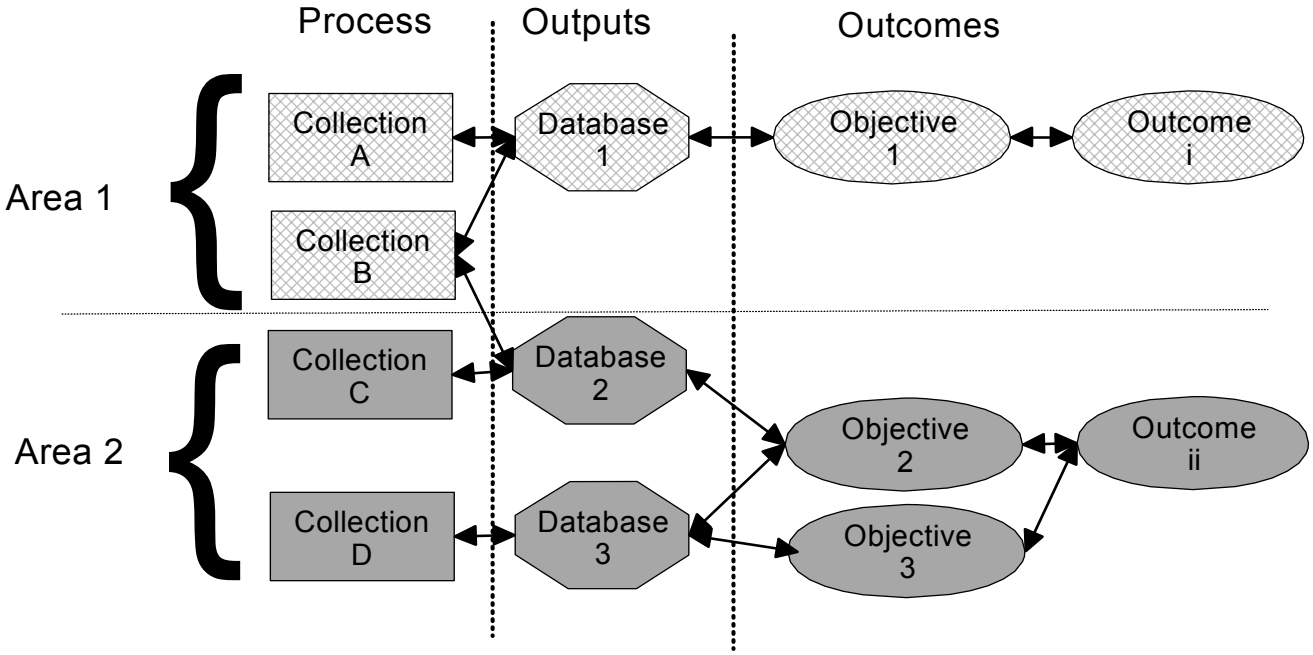
In the broad context of the ABS, the outcome might be considered to assist in informed decision-making, whereas the outputs might include deliverables such as statistics in publications and on the ABS Website, and consultancy devices. For policy makers, the outcomes would often correspond to 'real-world impacts', with the outputs corresponding to policy implementation and evaluation and the inputs corresponding to policy development.

In the context of quality-informed decisions, the underlying structure can be broadly interpreted as follows:

- the *inputs* correspond to the resources used in collecting data;
- the *process* corresponds to the data collection process;
- the *outputs* correspond to the data collected and the way the data are collated, stored and made available; and
- the *outcomes* correspond to the underlying reason the data are required.

This can be expanded further by recognising *objectives*, which relate specifically to the aims of a data collection. The relationships between processes, outputs and outcomes have been broadly summarised in *Figure 2* below:

Figure 2 - Processes, Outputs and Outcomes



In this example, there are two separate areas (Areas 1 and 2), responsible for specific objectives and outcomes. Area 1 manages Collection A and Collection B, while Area 2 manages Collection C and D, but also uses data from Collection B.

As shown in the above example, a database may access data from more than one data collection. Similarly, an objective may reference more than one database and outcomes may have multiple objectives. Also, some data collections will feed into more than one outcome. This reinforces the fact that data collections are not always custom-built for all needs and foreshadows the need for individual Quality Assessments.

The quality of the *processes* and *outputs* are measured using a data quality framework in the context of the *outcomes*.

4.2 The Data Quality Framework

The data quality framework proposed for incorporation into the decision cycle is based on a framework developed by Statistics Canada, which identifies six key dimensions of data quality:

- Relevance
- Accuracy
- Timeliness
- Accessibility
- Interpretability
- Coherence

This data quality framework has been published internationally (Brackstone G., *Managing Data Quality in a Statistical Agency*, (1999) Survey Methodology, Vol. 25, no. 2, Statistics Canada) and has been recommended by the ANAO as 'better practice' in specifying performance measures (*ATO Performance Reporting under the Outcomes and Outputs Framework, Australian Taxation Office*, Audit Report No.46 2000-01, pp63-64.) on advice from the ASB Statistical Consultancy Unit.

More specifically, the six dimensions of quality can be described as follows:

Relevance - The ***relevance*** of statistical information reflects the degree to which it meets the real needs of clients. It is concerned with whether the available information sheds light on the issues most important to users. Relevance is generally described in terms of key user needs, key concepts and classifications used and the scope of the collection (including the reference period). These components are then compared against specific user needs to assess relevance.

Accuracy - The ***accuracy*** of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterised in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of major sources of error that potentially cause inaccuracy (e.g. sampling, non-response).

Timeliness - The ***timeliness*** of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available.

Accessibility - The ***accessibility*** of statistical information refers to the ease with which it can be referenced by users. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of *accessibility* for some users.

Interpretability - The ***interpretability*** of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilise it appropriately. This information normally covers the availability and clarity of metadata, including concepts, classifications and measures of accuracy. In addition, interpretability includes the appropriate presentation of data such that it aids in the correct interpretation of the data.

Coherence - The ***coherence*** of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. Coherence encompasses the internal consistency of a collection as well as its comparability both over time and with other data sources. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys.

4.3 Framework for defining data needs

The process for defining a data need can be broadly summarised as describing the data need in terms of what an ideal data source might look like. The ITO model is used to help ensure that the data need remains focused on the underlying decision that needs to be made (i.e. the outcome), while the data quality framework is used to help clarify the data need by ensuring that the data need considers all the aspects of quality.

Table 1, below, provides a list of typical issues that need to be considered in defining a data need. This clarification of the data need indicates what is considered 'fit for purpose' and sets the standard against which data sources can be evaluated.

Table 1 - Typical issues to consider when framing a data need

Dimension	***Examples of Data Need Requirements***
Relevance	● How will the data be used in the decision-making process? ● What concepts do we need to measure? ● What population are we interested in? ● What classifications are we interested in?
Accuracy	● What are our sampling error requirements? ● What level of estimates are required?
Timeliness	● How soon do I need the data? ● How recent does the data need to be?
Accessibility	● How are the data made available? ● Will the cost of the data be prohibitive? ● In what forms are the data available (unit record file versus aggregates, electronic versus hardcopy)?
Interpretability	● Are the data sources documented sufficiently to known that the data source matches my needs?
Coherence	● What comparisons will be required over time? ● What comparisons will be required with other data sources? ● Is it important to match with certain standards?

4.4 Role of ABS

The ABS has a key role in the defining of data needs through its leadership of the National Statistical Service. This role can be broadly summarised as assisting users of data (i.e. government and the community) to define their data needs for:

- the development of new data collections; and
- the assessment of existing data sources.

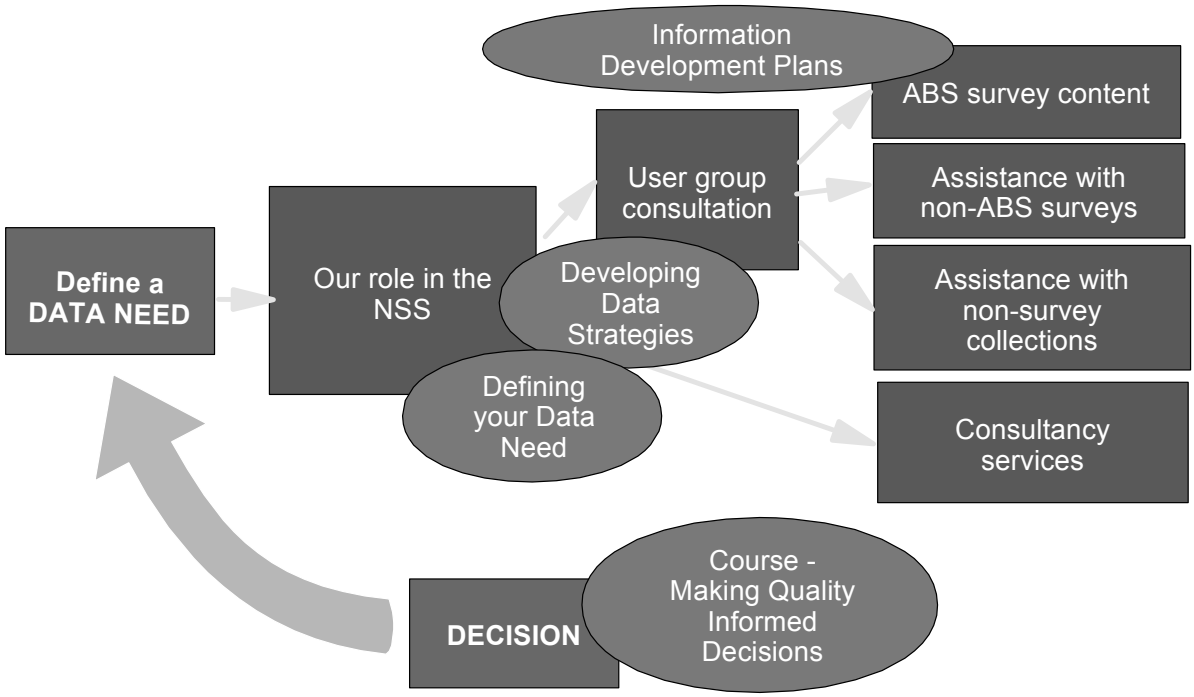
This assistance may be provided either through general user education, or specific involvement through mechanisms such as user groups, outposted officers or the provision of consultancies. Typical examples of specific involvement include:

- provision of internal and external seminars and training courses (e.g. the recently

piloted 'Making Quality-Informed Decisions' or courses on defining data needs or developing data strategies);

- user consultation on the content of ABS surveys;
- direct assistance to other government agencies with non-ABS surveys (e.g. methodological or subject matter assistance); and
- direct assistance to other government agencies with non-ABS administrative collections.

Figure 3 - The Role of the ABS in Defining Data Needs



5. Describing Data Sources

The purpose of describing the quality of data sources is to provide people with sufficient information to assess the 'fitness for purpose' of the data source against their specified data need. This description of the quality is called a Quality Declaration.

5.1 The Quality Declaration

The Quality Declaration focuses on the *process* of the data collection and the data *outputs* by providing a primarily descriptive overview (i.e. *qualitative measures*) of the data collection, supported by a small number of performance indicators (i.e. *quantitative measures*) for those characteristics where it is appropriate (e.g response rates).

For each data collection, the Quality Declaration provides information about the collection's methodology and processes. The Quality Declaration provides the necessary background for primary and secondary users (whose knowledge about

the data collection may be restricted to the information contained within the Quality Declaration) to complete an informed Quality Assessment (as described in **Risk Management**). The Quality Declaration will also be useful as a general reference document, for those needing to gain a broad understanding of the data collection.

It is the role of the Data Collection Managers to complete Quality Declarations for each of their collections. However, there may be data collections referenced within ABS for which there is no Data Collection Manager, primarily because the responsibility for managing the data collection exists fully outside ABS (e.g. managed by another government department). In this case, two main options exist:

- gain the cooperation of the area responsible to complete a Quality Declaration; or
- allocate the responsibility for completing the Quality Declaration to an area within ABS (such as the primary user of the data from the data collection). In this case, it is likely that the area may still need to seek advice from the area rsponsibile for managing the data collection.

5.1.1 Defining the content of the Quality Declaration

A draft template has been developed to assist areas to complete Quality Declarations and can be found in **APPENDIX 1 Template - Quality Declaration**. The Quality Declaration template draws upon the data quality framework described above in **4.2**.

The template identifies a range of characteristics for each of the six dimensions of data quality. These characteristics provide an overview of the associated quality-related issues. For example, the characteristics selected to represent 'accuracy' include the level of sampling error, the response rate, adjustments to data, levels of training and comparability in data values with related data sources. For each of these characteristics, examples of typical qualitative and/or quantitative measures have been provided.

The characteristics included in the template represent a first draft of what could be included in a Quality Declaration. However, the content is still very much open for debate. For example, it might be argued that some of the issues surrounding data security need not be there, while timeliness would benefit from including the timing of the first publication from the data collection and the release (where applicable) of a confidentialised unit record file.

The final content should be agreed by PSG Data Management and Dissemination, ESG Data Management, IMD Data Management and Statistical Consultancy and Training within MD.

5.1.2 Completing in the Quality Declaration

The Quality Declaration should be completed to allow an informed Quality Assessment to be completed. This is not to say, however, that the Quality Declaration needs to be a lengthy document which requires excessive effort to complete and maintain. Rather, in keeping with the concept of quality as being 'fit

for purpose', the Quality Declaration simply needs to be sufficient for users to be able to assess the appropriateness of a data collection for their own requirements.

For each item, only a short qualitative description (one to two paragraphs in most instances), a response to a list of choices or the provision of some quantitative information should be sufficient.

It is recommended that the person completing the Quality Declaration be familiar with the completion and use of the Data Collection Assessments.

For this purpose, **APPENDIX 1 Template - Quality Declaration** includes:

- a descriptive overview of each dimension of quality;
- the data characteristic;
- the questions to be answered by the Collection Manager;
- a definition / explanation of the data characteristic; and
- an explanation of how the information contained in the Quality Declaration might be used in a Quality Assessment.

5.2 Role of the ABS

The ABS has a key role in describing existing data sources, both as a major disseminator of statistics and as a leader for the National Statistical Service.

As a leader, the ABS has a responsibility for providing advice on the content of the Quality Declaration, reaching corporate agreement on a standard template to be used by all ABS data collections (except possibly derived collections as the Quality Declaration is better suited to describe data sources one at a time). As noted above, the final content of the Quality Declaration should be approved by PSG Data Management and Dissemination, ESG Data Management, IMD Data Management and Statistical Consultancy and Training within MD.

A further extension of this leadership role is to disseminate guidelines and templates to the broader National Statistical Service, so that other data providers are also able to document their data collections using Quality Declarations. This approach has already been initiated with the ABS Statistical Consultancy Unit providing advice to the Department of Education, Science and Training on the development of a Data Collection Assessment Framework. Also, as with defining data needs, these methods need to be documented with appropriate learning vehicles (e.g. reference documents, on-line learning, seminars or formal courses). Such strategies should be coordinated where possible with international efforts by government statistical agencies in the field of quality.

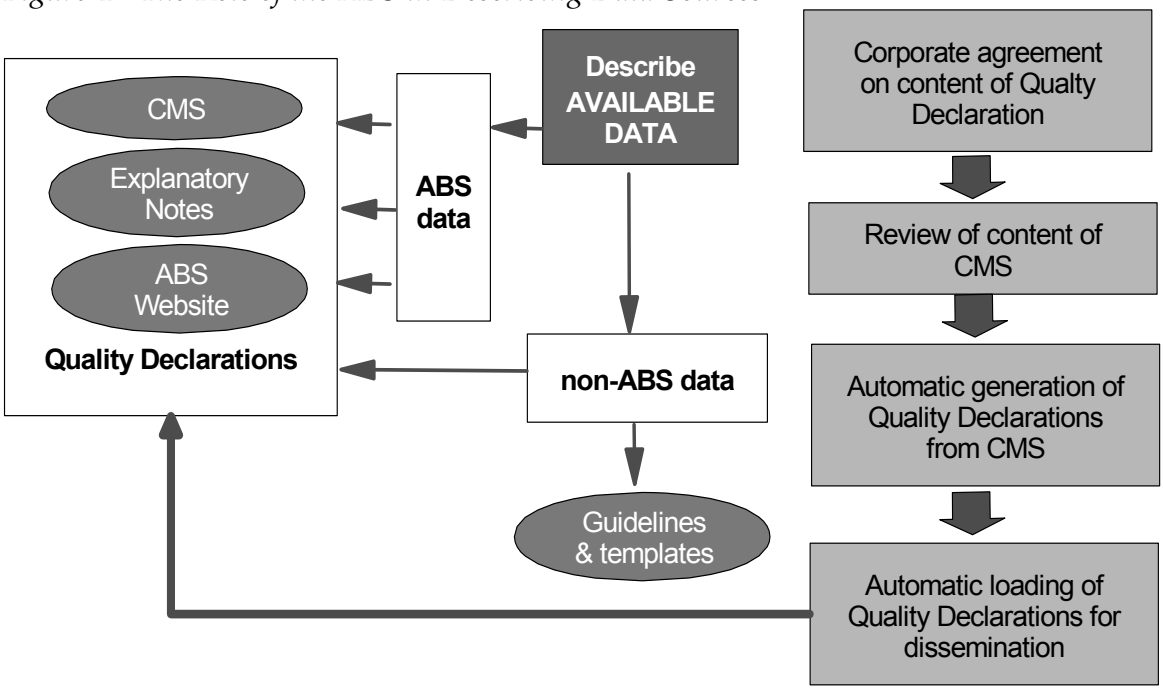
As a major disseminator of statistics, ABS also has a role in disseminating completed Quality Declarations through avenues such as the ABS Website, the Directory of Statistical Sources and ABS publications. In ABS publications, the Quality Declaration forms a natural base for Explanatory Notes, providing the Explanatory Notes with both a consistent structure across all ABS collections and a basic

minimum information to be included.

However, to be able to readily disseminate Quality Declarations the ABS must first agree on the content of the Quality Declarations. Then, as the corporate repository for metadata, the Collection Management System (CMS) will need to be reviewed in terms of its content and how well its fields will provide the necessary information required for the Quality Declarations. Next, it is desirable for the Quality Declarations to be generated automatically from the Collection Management System. This process has already been tested, with a successful prototype having already been developed.

Finally, it would also be desirable to facilitate automatic loading of the Quality Declarations for dissemination into ABS publications (through PPW) and to the ABS Website.

Figure 4 - The Role of the ABS in Describing Data Sources

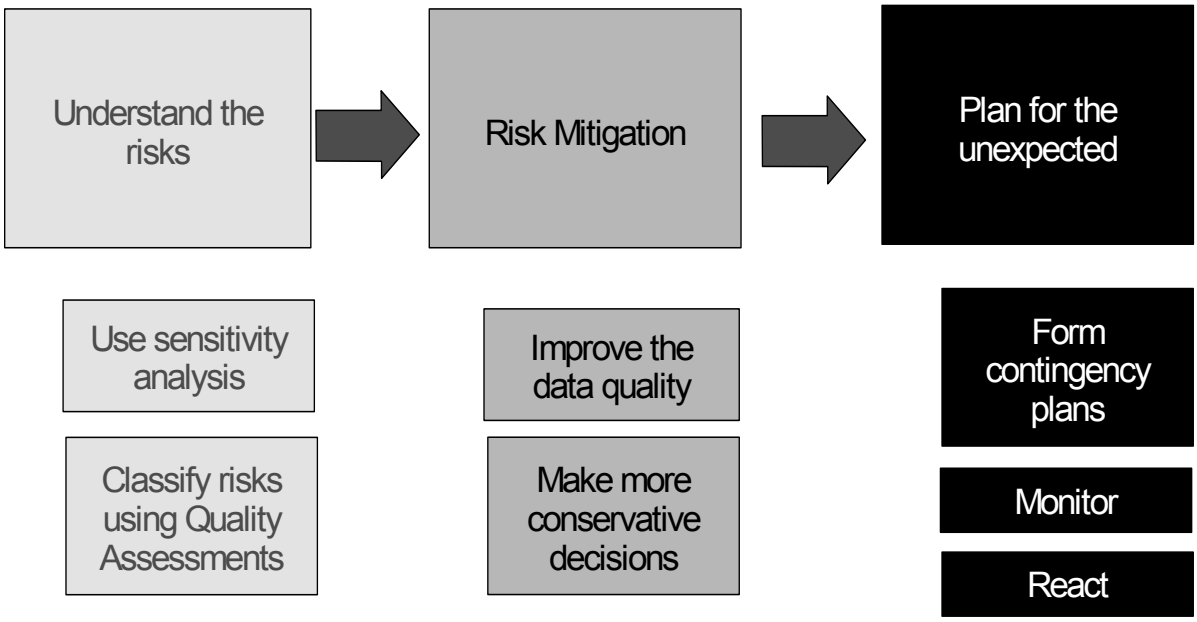


6. Risk Management

Risk management is the process by which information about the quality of the data is integrated into the decision-making process. If the data quality is poor, then the risks of making a poor decision using those data are greater. Conversely, if the data quality is high, then greater confidence can be placed in the information being used to make an informed decision. Thus, the concept behind making quality-informed decisions is to make decisions where the risks of using the data are appropriately managed.

For the purposes on facilitating quality-informed decisions, risk management has been subdivided into three stages which are broadly summarised in *Figure 5 - Risk Management* below:

Figure 5 - Risk Management



6.1 Understanding the risks

The first step in risk management is to understand the risks, and recognises two key steps: sensitivity analysis; and classifying risks using Quality Assessments.

The purpose of sensitivity analysis to identify the various levels of risk associated with using a specific data source for a given data need. This is achieved by examining each of the data characteristics identified in the Quality Declaration and trying to understand the potential impact on the underlying decision of using the data.

As a result of this analysis, each characteristic is classified according to the degree of match between the data need and the data source, ranging from *'the data collection significantly falls short of requirements'* to *'the data collection significantly exceeds requirements'*. This process is documented in a Quality Assessment form.

6.1.1 Sensitivity Analysis

Sensitivity analysis is best explained as the process of considering possible alternate scenarios and their potential impact on the underlying decision. In most cases, this will focus more specifically on considering how the data values might vary across a range of different scenarios and how the different values might lead to different decisions. This can be as simple as going through a process of asking two key questions:

- How different would the data need to be for me to make a different decision?
- How likely is it that the data would be that different as a result of the mismatch between my data need and the data source?

Typical examples would be the impact of sampling error, response rates and mismatches in scope and classifications between the data need and the data source. However, scenarios might also consider issues such as limitations of the type of analysis that can be performed (if there are problems with accessibility, for example).

These scenarios are often, by nature, subjective. For example, for response rates, different scenarios are created based on a series of subjective judgements on how non-respondents might differ from respondents and the likelihood of such a difference occurring. Where possible, additional information should be used to verify these assumptions (such as analyses of non-respondents).

The use of sensitivity analysis and how the different scenarios might impact on the underlying decision is discussed in greater detail in **APPENDIX 1 Template - Quality Declaration** which includes an explanation of how the information contained in the Quality Declaration might be used in a Quality Assessment.

6.1.2 Classify Risks Using Quality Assessments

The second part of understanding the risks is to classify the risks using a Quality Assessment template. The Quality Assessment follows the same format as the Quality Declaration by addressing each of the characteristics within each dimension of data quality, but requires a simple subjective assessment as to whether the specific characteristic for that data collection meets the user's specific requirements (with a short explanation). Thus, each characteristic are classified into one of five categories accoridng to the associated level of risk:

1. *the data collection significantly falls short of requirements;*
2. *the data collection is sufficient with some areas of reservations;*
3. *the data collection is sufficient for the requirements;*
4. *the data collection significantly exceeds requirements; and*
5. *there is insufficient information to judge the suitability of this characteristic*.

In addition, the Quality Assessment requires an indication of the overall suitability of the data collection for the user's requirements.

The purpose of the Quality Assessment is to identify how well the data collection *process* and data *outputs* meet the needs of the user. This includes identifying both limitations (which should impact on the way the data are used) and instances where the data collection exceeds the user requirements (indicating potential areas for savings with regards to that particular user need).

It is the role of the key users or clients of the data collections to complete Quality Assessments and is suitable for assessments from people both internal and external to the ABS.

These assessments can play a key role in assessing the appropriateness of a given collection in meeting the data need associated with a particular objective and can be

used as a basis for improvements to the data collection or making appropriate risk management strategies when the data needs are not perfectly met (as described in **6.2 Mitigating risks**). Thus, it is also in the best interest of the user to complete a Quality Assessment.

For each data collection there would be a separate Quality Assessment for each different objective (noting that this includes both single users with multiple objectives and multiple users each with a single, but different, objective).

A template has been developed to assist areas complete Quality Assessments and can be found at **APPENDIX 2 Template - Quality Assessment**.

6.1.2 Completing in the Quality Assessment

As the Quality Assessment helps identify how well the data meet the needs of the user, it needs to be filled out from the perspective of how the data quality of each characteristic impacts on the way the data can be used (using the information in the Quality Declarations).

These assessments are of the form:

With regards to this characteristic,

- ❑ *the data collection significantly falls short of requirements;*
- ❑ *the data collection is sufficient with some areas of reservations;*
- ❑ *the data collection is sufficient for the requirements;*
- ❑ *the data collection significantly exceeds requirements; or*
- ❑ *there is insufficient information to judge the suitability of this characteristic.*

Comments:

_____
_____
_____
_____
_____

The *Comments* field is used to explain and document why the particular assessment was made. The type of explanation should refer to the user's own requirements and note specific reasons why the characteristic might be considered deficient or exceeding their requirements (or what additional information would be required to be able to make an assessment).

Note, however, that an assessment is not required for all characteristics included in the Quality Declaration. An explanation of why certain characteristics have been excluded from the Quality Assessment has been provided in the template at **APPENDIX 1**.

In addition, a general summary of the Quality Assessment is provided at the start of

each Quality Assessment. This overall assessment takes the form:

In general this data collection:

- ❑ *significantly falls short of requirements for addressing the outcomes;*
- ❑ *is sufficient with some areas of reservations;*
- ❑ *is sufficient for addressing the outcomes; or*
- ❑ *significantly exceeds requirements for addressing the outcomes.*

The main shortcomings are in the characteristics of:

_____
_____
_____
_____
_____
_____
_____
_____

<<provide bullet point list of characteristics rated as significantly falling short of requirements>>.

The main strengths are in the characteristics of:

_____
_____
_____
_____
_____
_____
_____
_____

<<provide bullet point list of characteristics rated as significantly exceeding requirements>>.

Finally, having completed a Quality Assessment, it can be quite useful to transfer the results to a Quality Assessment Summary Table as depicted below in Table 2 below. In the table, each data characteristic is recorded in a cell corresponding to its appropriate dimension of quality and the Quality Assessment it received. For example, if the response rate fell short of the requirements, then it would be recorded in the row 'Accuracy' and under the column heading 'Falls short of requirements'. This table provides a quick overview of the identified risks for using the data collection for the specified data need.

Table 2 - Quality Assessment Summary Table

	Insufficient information	***Falls short of requirements***	***Sufficient with some areas of reservation***	***Sufficient***	***Significantly exceeds requirements***
Relevance					
Accuracy					
Timeliness					
Accessibility					
Interpretability					
Coherence					

6.2 Mitigating risks

In **6.1 Understanding the risks**, risks were identified by comparing the data need to the data source, as documented in a Quality Declaration. The potential impact of these risks were understood through the application of sensitivity analysis and the results were documented in a Quality Assessment.

The next step is to investigate options for reducing the level of risk. This framework has identified two such avenues:

- Improving the data quality; and
- Making more conservative decisions.

6.2.1 Improving the data quality

Improving the data quality primarily deals with looking for opportunities to improve the match between the data need and the data source. These opportunities can be divided into those which involve directly modifying the data collection and those which do not.

Directly modifying the data collection will generally only be an option where the user is identified and accepted as a major stakeholder. Even then, there is a need to balance the needs of one user or a group of users across the needs of all users, noting different decisions will lead to different data needs and therefore different ideal data sources. Also, the degree to which a data collection can be modified and the effort and resources required for such modifications will vary considerably for different data collections. Any decision to modify the data collection should be the result of a cost-benefit analysis with the cost of implementing changes assessed against the benefits

Modifications to a data collection may stem from any of the Data Collection Assessment ratings for characteristics:

- *The data collection significantly falls short of requirements.*

  The modification should be structured so that the characteristic in question will better meet requirements. For example, the scope of the data collection might be

changed to include a broader geographical scope, the sample size might be increased to meet specific user requirements, or the databases might be improved to enable easier access to the data.

Naturally, any such modifications would first need to be fully costed and a decision made as to whether the improvement is warranted. Note, however, that given the characteristic falls significantly short of requirements, it is likely that the characteristic is significantly compromising the quality of the data and potentially the decisions being made on the basis of these data.

- *The data collection is sufficient with some areas of reservations.*

  These modifications will be similar to those above, although the assessment has already identified that the data collection is already nearly sufficient to meet needs. As such, it is likely that efforts will be focused on the more crucial areas where characteristics have fallen significantly short of requirements. However, it may also be that the costs of modifying the data collection sufficiently is comparatively small.

- *The data collection is sufficient for the requirements.*

  While the data collection may meet the specific needs of the user, the user may have identified some potential areas for improvement, which may involve only marginal additional costs or allow for the data to be used more effectively across a wider range of purposes.

- *The data collection significantly exceeds requirements.*

  In these instances, it appears likely some savings might be achieved by reducing what is being offered. For example, the amount of editing might be reduced, it might be decided that a lesser degree of documentation would suffice, or it might be decided that there are insufficient benefits in preserving an outdated classification.

- *There is insufficient information to judge the suitability of this characteristic.*

  This reflects a need to improve the level of documentation. Without the information necessary to make an assessment of the characteristic, it is unknown whether the data will meet needs. In addition to exposing users to making incorrect inferences, it is not possible to identify where improvements or savings can be targeted.

There are also a number of options for reducing the level of risk which do not involve directly modifying the data collection:

- *apply a data collection to only part of the problem.*

  This option accepts that the data source is sufficient for answering part of the

question, but not all of it. For example, the scope of the data collection might only cover specific States but information is required for all States and Territories. Alternatively, only certain information might be accessible meaning that only part of the question can be addressed. In such cases, it might be preferable to only use the data to address only part of the question, particularly when the risks associated with making inferences about the areas not covered by the data source are high. This option combines well with the next option of accessing multiple data sources.

- *accessing multiple data sources.*

  Accessing multiple data sources provides two benefits. Firstly, it provides an opportunity to check for external validation of the data sources through looking for consistency in the data values across different data sources (in the context of their respective levels of quality).

  Secondly, accessing multiple data sources opens the door for using each data source for its respective strengths and using other data sources to cover the weaknesses. This might be as simple as using different data sources for different States and Territories when a national picture is desired. Alternatively, multiple data sources could, for example, be used in synthetic estimation to provide small area estimates. One example of this would be in producing small area estimates for people with disabilities where propensities for disabilities across various demographics can be estimated using survey results which can then be applied to small area data from the Census of Population and Housing.

  Another example would be to use additional data sources to test or refine the assumptions used in the sensitivity analysis. For example, a more frequent survey might be used to estimate trends as a basis for developing scenarios on how much data may have changed since the data collection was last run. Similarly, the confidence intervals for comparable estimates from different data sources could be compared to restrict the likely range of values the true population value might take.

- *deciding more information on the data collection is required.*

  This option specifically addresses the Quality Assessment rating of '*There is insufficient information to judge the suitability of this characteristic*' by suggesting that more research is done to provide sufficient information to assess the suitability of the characteristic in question. Without sufficient information, the risks associated with the data characteristic are high. For example, it might not be possible to know whether the risk of a high level of non-response bias is high as the response rate is not known.

- *deciding a new data source is required.*

  Finally, it may be decided that there is no data source or group of data sources which can be used to adequately feed into the underlying question - the risks

associated with existing data sources are too high and cannot be sufficiently mitigated. In this case, it might be necessary to develop a new data source.

If a decision is made to investigate options associated with developing a new data source, the same process of assessing the proposed data source against the specified data need using the data quality framework should be followed.

It should also be noted that often there are insufficient resources to address all aspects of quality perfectly, so compromises need to be made to achieve an 'affordable level of quality'. However, in making these compromises, two issues need to be considered:

- Where should compromises be made?
- Once compromises have been made, will the data still meet data requirements sufficiently?

6.2.2 Making more conservative decisions

Having identified the risks, it is important that the underlying decision takes these risks into account. In other words, these decisions should take into account the quality of the data as well as the values of the data. As such, making more conservative decisions is specifically aimed at the person making the decision which has generated the data need.

It is difficult to provide specific options here, as the options are dependent on the underlying decision and the corresponding areas of risk. However, it is important to understand that these options do exist. One example might be a decision on whether the allocated budget will support subsidising a new drug as part of the Pharmaceutical Benefits Scheme where more conservative decisions might include implementing localised trials, reducing subsidy levels at first, restricting eligibility criteria for receiving the subsidy (e.g. need a Seniors Card), delaying a decision pending more information or even deciding to use the money to subsidise a different drug where expected usage patterns are better understood.

6.3 Planning for the Unexpected

In the first step, the risks were identified. Then, in the second step, attempts were made to mitigate these risks. In this final step, plans are developed and put in place in case these risks are realised:

- *form contingency plans* in case the identified risks are realised;
- *monitor* what is happening in case the risks are realised; and
- *react* to the realised risks when they are identified, using the prepared contingency plans.

6.3.1 Form contingency plans

In understanding the risks, the Quality Assessment template was used to classify the

risks. The areas which were identified as higher risks (usually classified as significantly falling short of requirements) are the same areas where contingency plans are required (unless the risks were later mitigated sufficiently).

Contingency plans are simply strategies of what to do if certain risks are realised. For example, a low response rate for a survey generates a risk that the survey results are significantly influenced by non-response bias. As a result, inappropriate decisions might be made on the basis of the biased results. This is covered to some degree in sensitivity analysis, but a subjective judgement is still made on the likely degree of non-response bias. Thus, it is important to have a plan in place if later information suggests that earlier decisions were inappropriate. Continuing with the example of the low survey response rate, it would be wise to have a contingency plan in case later information did suggest that the survey had suffered from a higher than expected degree of non-response bias.

These contingency plans should relate to the underlying decision. Using the PBS example referenced earlier on basing funding a subsidy for a drug on existing usage levels, a contingency plan might be to reduce or drop the subsidy if usage levels turn out to be much higher than expected. Similarly, the eligibility criteria for receiving the subsidy could be restricted. These plans can be very similar to those considered early at the risk mitigation stage. However, instead of mitigating the risk immediately through making a more conservative decision, the decision might not fully take into account the associated risks. Rather, the risk mitigation option would only be implemented if further information suggested that the risks had been realised.

6.3.2 Monitor and React

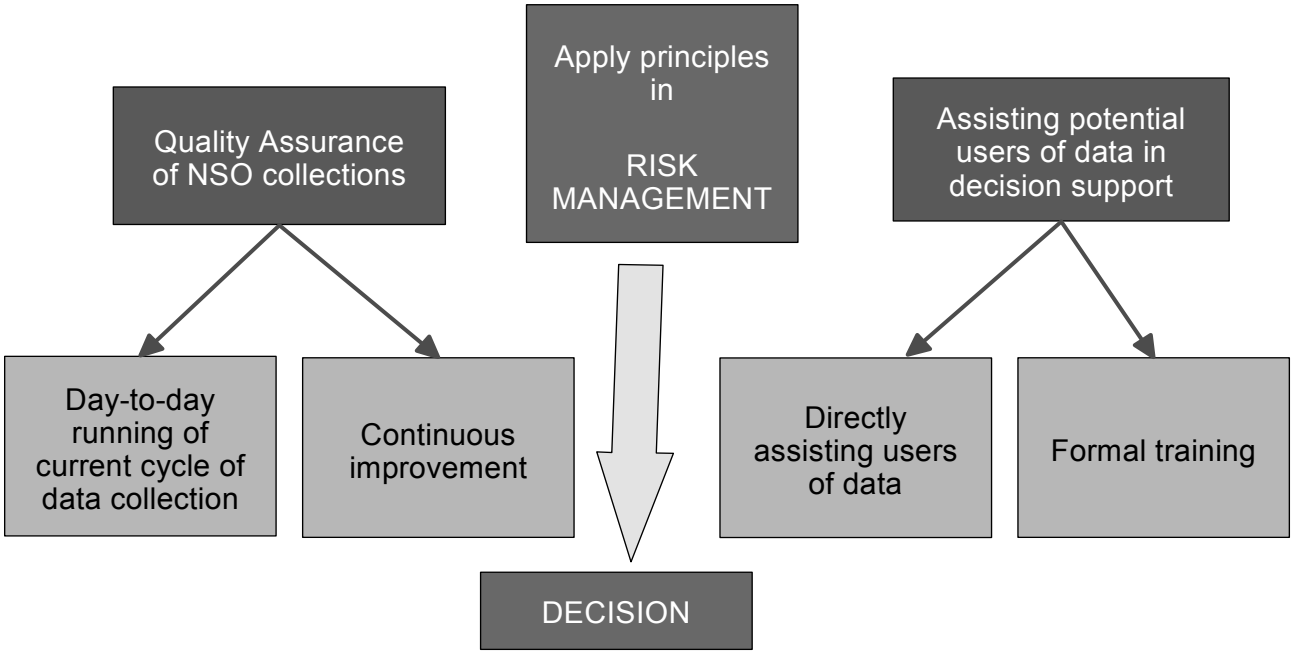
Monitoring is a key part of planning for the unexpected. Having formed contingency plans, it is important that the information is available which will trigger these contingency plans into action.

While it may be possible to continue to monitor data from a regular survey or an ongoing administrative collection, this will not always be possible. As such, it is important to also consider other ways to monitor the impact on the underlying decision. For example, monitoring budgets would assist in avoiding overspending budget allocations. Similarly, a decision to run specialised training programs for the unemployed would benefit from monitoring both participation levels in training programs, participant comments on the training and overall levels of unemployment.

6.4 Role of the ABS

The ABS has two specific roles in risk management, which can be broadly described as quality assurance of ABS collections and 'decision-support' for users of data.

Figure 6 - The Role of the ABS in Risk Management



Quality assurance of ABS collections is an area where the ABS has much experience. With the benefit of the decision cycle and the data quality framework, it is hoped that ABS efforts in this area can be further improved. Quality assurance includes two distinct aspects - the close day-to-day scrutiny of a current cycle as it progresses to survey clearance, ensuring that the data quality is sufficient to be published, and the broader review process which considers where efforts should be focused in the spirit of continuous improvement.

Through continued efforts of developing appropriate quality measures, ABS will be better placed to monitor the data quality and respond accordingly. However, just as data need to be relevant, accurate, timely, accessible, interpretable and coherent, so do the quality measures. Thus the quality measures must be relevant, feeding directly into our own decision cycle. Similarly, automated generation, loading and presentation systems for quality measures will make the quality measures more timely and accessible. Appropriately documenting and presenting the quality measures will aid in their interpretability. In keeping with those principles, it is important that, as the data collection progresses through data processing and estimation, these measures are readily available and continuously updated so that those monitoring the collection are able to respond quickly to problems with the data.

The data quality framework is also valuable for more focused methodological reviews of data and their associated collections. Areas for improvement can be identified using processes comparable to the Quality Assessment process described earlier, with efforts being focused at improving those errors where the greatest (affordable) gains in quality can be realised, or conversly how quality can be best maintained under reduced resourcing.

The second role available to the ABS is that of assisting the real users of the data in 'decision support'. Decision support can be described as assisting users to make their Quality Assessments and then apply the results of these Quality Assessments to their underlying decision through the adoption of appropriate risk management strategies. This can be achieved either through directly assisting the data user (e.g. user group consultations, outpostings, consultancies) or through formal education (e.g. training courses, seminars, on-line learning, reference materials).

APPENDIX 1 Template - Quality Declaration

The template presents the six dimensions of data quality (as described in the main report in **4.2**):

- Relevance;
- Accuracy;
- Timeliness;
- Accessibility;
- Interpretability; and
- Coherence.

For each dimension, the template gives a number of relevant data characteristics which need to be assessed. For example, for 'relevance', some of the characteristics are the scope, reporting unit, frame, classifications and concepts. A definition for each characteristic is provided in the third column. In the last column, there is an explanation of how that characteristic is important from a user assessment perspective. For example, if the 'scope' of the collection excludes people or groups that the user is interested in, the user will need to make some judgements about how these exclusions impact on their decision making capability.

Relevance

The ***relevance*** of statistical information reflects the degree to which it meets the real needs of clients. This is addressed in the Quality Assessment by:

- looking for mismatches in scope, classifications, concepts and data items between what the data collection provides and what the user requires; and
- understanding who the respondents are and how the information is collected.

Looking for mismatches is important because a mismatch tells us that the data collection is not measuring exactly what user wants to measure. As such, it is important to understand the potential impact of the mismatch on the decisions that the user wishes to make.

Understanding who the respondents are (e.g. universities) and how the information is collected (e.g. electronically via e-mail or the Web) is important because this assists in better understanding the limitations of the resulting data.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Scope	What is the population <u>actually</u> covered by the data collection?	Scope includes both the geography covered by the data collection (e.g Victoria) and any other rules used to identify whether a unit is included or not (e.g. exclude people 15 years or under, exclude non-residents) .	If the population covered by the data collection is different to the population you are interested in, then the following questions need to be asked: ● For parts of the population that you are interested in, but are <u>not</u> available in the data collection, are they likely to exhibit different characteristics? ● Can you subset out parts of the population that you are <u>not</u> interested in, but are included in the data collection? If not, are these additional units likely to exhibit different characteristics? In answering these questions, it is also important to remember the impact on totals as well as averages. For example, missing out on a part of the population is likely to mean that the totals will be too low.
Reporting Unit	Who collates and provides the data?	The reporting unit describes who actually provides the data. In some cases, the reporting unit will also be the unit of interest. However, this will not always be the case (e.g. universities might report course information on students). In those cases where the data are provided and collated by different people, please details of both.	The reporting unit is important as the information collected will generally be from the perspective of the reporting unit. For example, collecting information on fields of study from students and the institution they are studying at may well produce different results.
Frame	How is the list of potential respondents compiled? Are there any data quality issues associated with the frame (e.g. new units, defunct units, duplicates, age of frame)?	The results from a data collection are highly dependent on the list used to identify who should respond to the data collection. The quality of this list will have a strong impact on the quality of the data.	For example, a list prepared using the White Pages would exclude households without home telephones or with silent numbers. The issues here are similar to those identified for scope.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Classifications used	What are the key classifications used?	A classification is set of defined groupings or categories - based on common relationships - into which all members of statistical units can be divided or arranged. These groupings or categories can be ordered systematically, are mutually exclusive and exhaustive, and are based on one or more data items. Examples of classifications include: State, Industry; Highest Level of Educational Attainment; Age (in 5 year groupings); and Country. In those instances where classifications used correspond to industry, national or international standards, this should be indicated.	If the classifications used in the data collection do not match up with requirements, then it is important to consider the potential impact of this. For example, a difference in industry classifications may mean that you are unable to exactly measure the industries you are interested in. As with scope, it is important then to assess the likely impact of this mismatch.
Concepts used	Describe any key concepts addressed in the data collection.	A concept in the context of a data collection usually refers to an issue which is often difficult to measure directly (e.g. well-being, some economic concepts) or needs to be derived through several data items (e.g. unemployment, disability). Often the key concepts are the key issues which the primary user is seeking to measure in the data collection.	If the data collection is not measuring the exact concept you are interested in, it will be necessary to assume that the concept you are interested in would produce similar results to those in the data collection, had it been measured. The greater the difference in the concepts, the more tenuous this assumption becomes and the greater the danger that decisions will be made using data which are not conceptually relevant to your needs.
Key data items	What are the key data items collected?	A data item is a particular characteristic which is measured or observed. There are two main types of data items: ● Parametric data items are quantitative measures and have both an associated unit of quantity (e.g. $, hectares, hours) and an associated type (e.g. flow, stock, index, movement). ● Classificatory data items are described in terms of a category (e.g. industry, state, country of birth) rather than using a quantitative or numerical measure.	For the collection to be useful, it needs to collect the information you are interested in. Mismatches in data items will lead to similar problems as mismatches in concepts or classifications.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Mode of data collection	What mode of data collection is used?	The mode of the data collection describes the method used to collect data. Examples include: ● e-mail; ● web; ● Computer Assisted Telephone or Personal Interview; and ● Personal Interview.	The way the data are collected may lead to certain limitations in the data, often relating to the scope of the data collection or the type of information that can be collected using that mode (e.g. personal interviews may cause problems with sensitive questions, but allow the interviewer to better clarify issues with the respondent).
Intended audience and purpose	For what purpose(s) is the data collection run? Who is the primary intended audience for the data?	The intended audience are the primary users of the data collection. In most cases, the data collection will have been designed specifically to meet the needs of these users. The purpose of the data collection is defined as the primary use for which the data will be used by the intended audience.	While a Quality Assessment is not required for this characteristic, this assists in providing the user with an understanding of the broader context of the data collection.
Owner of data collection	Who is responsible for managing the data collection? Who is responsible for deciding on the data items collected?	The data collection manager is defined as the person or position responsible for the operation coordination and running of the operational aspects of the data collection. In addition, a person or group of people will be responsible for deciding which data items are collected or included. This may differ from the data collection manager.	Once again, a Quality Assessment is not required for this characteristic.
Authority	Under what (and whose) authority and/or legislation is the data collection run?	This provides information relating to expected response rates and the general context under which the respondent is required to provide the data. For example, the quality of information collected under an Act of Parliament for the provision of federal funds might be expected to differ from that collected from university administrative records provided on a purely volunteer basis.	Not assessed in Quality Assessment.

Accuracy

The ***accuracy*** of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. As such, it is important to consider issues of both sampling error and non-sampling error (where applicable).

Issues such as mismatches in scope or classifications may also be considered here, but they are addressed primarily under Relevance.

For the Quality Assessment, the user needs to consider whether the accuracy of the data collection will be sufficient to meet their needs. If not, they then need to consider the impact of using the data. This may mean that decisions will be made using data from the data collection, when the underlying information that they are interested in could be significantly different. In other words, the data may be misleading, resulting in poor decisions.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Level of sampling error (applicable to survey samples only)	What are the relative standard errors for key data items? Include relative standard errors of key data items also for key subpopulations.	Sampling error reflects uncertainty in the true population value because information was collected from only a sample of the population. This is often measured as the relative standard error (i.e. standard error of the estimate as a percentage of the estimate). This can be used to identify a range of values that the true value is expected to lie between (e.g. 95% confidence interval).	If the range of values is high, this can impact on the decisions based on the data. For example, if you knew that the unemployment rate was in the range between 0% and 20%, would this restrict the type of policy decisions that you would be comfortable making?

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Response rates	What is the response rate? What steps are in place to attempt to maximise the response rate?	The response rate is calculated by dividing the number of responding units by the number of units which were selected and were in scope of the data collection. Examples of methods used to maximise the response rate include (but are not restricted to): • use of primary approach letters; • interviewers well trainined in establishing a rapport with respondents or the design of respondent-friendly questionnaires; • informing respondents how the results of the data collection will benefit them; and • detailed call back strategies.	In most instances, an assumption is made that the non-respondents would have provided similar information to the respondents. However, the non-respondents may in fact be quite different to the respondents, so the data will be biased to reflect those units which have responded. For example, imagine a data collection on university students where all the overseas students failed to respond. Had the overseas students responded, different conclusions may have been reached. In interpreting the response rate, it is important to consider how your conclusions based on the data may have changed if the non-respondents had responded very differently to the respondents. This is often best handled using a sensitivity analysis approach (see *6.1.1*). In completing the Quality Assessment, first consider how much the data would be likely to change and then consider how that might impact on any resulting conclusions or decisions made. Understanding the steps for maximising the response rate should provide some insight into the potential for non-repsonse bias.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Adjustments to data	What methods are in place for edits and data validation? What data items have more than 10% of units with missing values or have been edited or imputed? For imputed data items, approximately what percentage of units have been altered on the basis of editing or imputation? Are the data subject to large revisions?	Editing is the process of checking data records to ensure that they contain valid entries and changing the records where they do not, whereas imputation is the process of estimating data for individual records which have not been completed. Data validation is a general term for methods used to check that the data appear correct.	The concerns with high levels of editing and imputation is similar to the concerns associated with high levels of non-response. That is, how much are our decisions being influenced by data which didn't come directly from the respondents but were estimated? Similarly, if the data are subject to large revisions, there is a high degree of uncertainty about what the final data will actually be. Consider how much the data might change due to revisions and whether the revised data would lead to different decisions.
Other data issues	Are there any other issues that might impact significantly on the accuracy of the data?	Other issues may also impact on how well the data being collected actually measures what it is supposed to measure. Examples include: ● different levels of data quality for different data items in administrative collections; ● sensitive information; and ● recall bias.	Other issues, such as those listed here, can also influence the data collection's ability to accurately measure what the user actually wants to measure. For example, the respondent may not be able to provide the information with any degree of uncertainty, as they cannot remember the details sufficiently or they are being asked to provide an opinion on something on which they feel they do not have sufficient information. This information also needs to be considered as part of the accuracy of the data.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Level of training	What is the level of training received by those involved in the collection design and operation (i.e. questionnaire design, systems used to collect information, systems for editing and processing the data, etc.)?	Poor training can cause significant problems with the ultimate quality of the data. For example, questions could be misleading or ambiguous so the respondent may not have interpreted the questions as was originally intended. Similarly, poor training for data processing could lead to errors being introduced at data entry.	In making a Quality Assessment, the user needs to consider whether the level of training is sufficient for the data collected. This will be related to the nature and complexity of both the data collection procedures and the data to be collected.
Comparability in data values with related data sources	How does the data collected compare with similar data sources?	Comparability in data values with other data sources offers some insight into whether the data seem to be measuring what the user is interested in (noting that the user's requirements may be sufficiently different to prevent the use of the other data sources).	For this characteristic, the Quality Assessment focuses on whether a possible lack of comparability between the data values from this data collection and other related sources is sufficient to cause some concern with the data collection.

Timeliness

The ***timeliness*** of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Recency of data	What was the last reference period for the data collection? How often is the data collection usually run? When is the next data collection expected to occur?	The reference period refers to span of time to which the data refers. This may either refer to a single point in time or a span of time.	For this characteristic, the Quality Assessment is asking about the suitability of the timeliness of the data. If circumstances are likely to have changed significantly since the last time the data were collected (e.g. internet usage) and the data needs to reflect the current situation, there will be problems comparable to those experienced under the relevance and accuracy dimensions - the data may not be measuring what the user wants to measure which may lead to inappropriate decisions using that data. Thus, it may be concluded that the value of the data is limited given that the data are no longer relevant to the current situation.

Accessibility

The ***accessibility*** of statistical information refers to the ease with which it can be accessed by users. This is addressed in the Quality Assessment by considering:

- knowledge that the data exist;
- ease of accessing the data; and
- the security of the data. This aspect is for internal use only and should be removed from the DCO copy supplied to external clients.

This impacts on decisions regarding whether the data collection is an appropriate data source, with respect to ease of obtaining the data, its security and the impact on any dissemination of results.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Ease of getting data	What is the average time taken to fulfil a data request? What data are readily available on the Web? What publications are available, and where are those publications available? What is the associated pricing policy?	A data request will generally refer to a request for tabulated data. The complexity of the data request may vary for different requests so consider the average time required to meet a request of 'average' complexity. The pricing policy is the set of rules or guidelines for determining the cost for a user to purchase data.	Even having received permission to access, it might prove too difficult to get the data in a suitable form or it might take too long to get the data. Similarly, access to the data may prove to cost too much given your available resources.
Knowledge data exists	How are people internal to the department made available about the existence of the data? How are people internal to the department made available about the existence of the data?	Knowledge that the data exists is an important aspect of the accessibility of a data collection. This includes how information on the data collection is made available both internally (e.g. on the Collection Management System with the ABS or externally (e.g. on the Web or hardcopy publications in most libraries).	In the comments field, the user should indicate how they became aware of the data and how easy it was for them to locate the data. It is expected that ABS data are listed and documented on the Collection Management System.

Interpretability

The ***interpretability*** of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilise it appropriately. Interpretability has been addressed in the Quality Assessment by asking the questions:

- Is there sufficient information to make an informed Quality Assessment on all characteristics?
- How easy is it to obtain more information about the data and data collection if required?

If there is insufficient information to understand properly how well the data meets the user's specific needs, then they are in danger of using inappropriate and/or misleading data to make important decisions.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Level of documentation for data collection	Has this Quality Declaration been 'signed off' by the area in charge of the data collection? What more detailed information is available regarding the data collection?	The Quality Declaration is the document as described in this appendix. The level of documentation should be aimed at providing sufficient information for someone without previous knowledge of the data collection to complete a Quality Assessment (without using the assessment of 'insufficient information'). More detailed information might be available through other sources, such as user guides, a website or other documentation maintained by the data collection manager.	The Quality Assessment for this characteristic makes an assessment as to whether the level of documentation in the Quality Declaration is sufficient. Insufficient information to make a Quality Assessment means that there is uncertainty regarding the data quality for that characteristic. As such, any decisions using the data which are affected by that characteristic will be based on data on dubious quality and may lead to inappropriate decisions being made. The Quality Assessment should indicate that the level of documentation is insufficient for those characteristics which have been rated as "*there is insufficient information to judge the suitability of this characteristic* ". The comments field should indicate which characteristics have received this assessment.
Internal accessibility of documentation for data collection	Is the Quality Declaration ready available within the department?	The Quality Declaration should be available for all potential users of the data within the department to access, in case they need to review available data sources for a given need. Ideally the Quality Declaration should be stored on a corporately endorsed (standard) storage medium for documentation on data collections.	In the comments field, the user should indicate how easy it was for them to locate the Quality Declaration.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
External accessibility of documentation for data collection	Is this Quality Declaration available to people outside the Department (Web or Other - please specify)? Is more detailed information available on the Web (Web or Other - please specify)? What level of documentation is provided in publications?	This characteristic refers to the availability of the Quality Declaration to people who do not work within the department. More specifically, these people include anyone who might be interested in understanding the quality of the respective data collection (e.g. academics, policy analysts in other departments). Possible methods for external accessibility would include the inclusion of the Quality Declaration on a Website or in publications released to the general public.	For this characteristic, the Quality Assessment comments on whether documentation will be made sufficiently available for those outside the department. For users within the department, this assessment draws on whether it is important that people outside the department are able to access the documentation (e.g. to support published data). This would also alleviate the degree to which the area managing the data collection needs to be called upon to answer questions about the data collection.

Coherence

The ***coherence*** of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. This is captured in the Quality Assessment by focusing on changes over time to the data collection as any such changes will impact on any interpretation of how things may have changed over that period. For example, a perceived change in results between two time periods might simply reflect a change in definition. Thus, it is important to know when these definitions have changed and how much they have changed, and considering the potential impact of those definitional changes on the data.

Data Characteristic	***Questions to be answered by Collection Manager***	***Definition/Explanation***	***Relationship to Quality Assessment***
Consistency of classifications over time	List any changes in key classifications over time	To try and maintain or improve the general relevance of a classification, they are often reviewed and updated over time. Examples of classifications which are subject to review include: ● Statistical Local Areas (geographic); ● Collection Districts (geographic); ● Industry; and ● Countries.	The Quality Assessment should consider the changes in the key classifications in the specific context of the user's requirements. Some classifications may not be relevant to your needs or the changes may be minor compared to your needs. Alternatively, some changes may cause major problems in comparing data over time.
Consistency of concepts and methodology over time	List any changes in key concepts and methodology over time.	To try and maintain or improve the general relevance of a statistical concept, they are often reviewed and updated over time. For example, the concept of employment as measured may have changed over time. Similarly, other concepts such as innovation have evolved over time as more research is done in their respective fields. Similarly, changes in the collection methodology may impact on the resulting data. Examples might include changing the data collection methodology or the questionnaire.	The issues associated with this are the same as those listed above for assessing the consistency of classifications over time.

APPENDIX 2 Template - Quality Assessment

The data source being assessed is:

_____

The data need against which this data source is being assessed can be summarised as follows:

_____
_____
_____
_____
_____

In general this data collection:

- ❑ *significantly falls short of my requirements;*
- ❑ *is sufficient with some areas of reservations;*
- ❑ *is sufficient for addressing my data needs; or*
- ❑ *significantly exceeds my requirements.*

The main shortcomings are in the characteristics of:

_____
_____
_____
_____
_____
_____
_____
_____

<<provide bullet point list of characteristics rated as significantly falling short of requirements>>.

The main strengths are in the characteristics of:

_____
_____
_____
_____
_____
_____
_____
_____

<<provide bullet point list of characteristics rated as significantly exceeding requirements>>.

Relevance

Scope	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Reporting Unit	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Frame	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Classifications used	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Concepts used	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Key data items	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Mode of data collection	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Authority	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Accuracy

Level of sampling error	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Response rates	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Adjustments to data	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Other data issues	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Level of training	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Comparability in data values with related data sources	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Timeliness

Recency of data	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Accessibility

Ease of getting data	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Knowledge data exists	I*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Interpretability

Level of documentation for data collection	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Internal accessibility of documentation for data collection	I*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

External accessibility of documentation for data collection	I*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____

Coherence

Consistency of classifications over time	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____
Consistency of concepts over time	*With regards to this characteristic,* ❑ *the data collection significantly falls short of requirements;* ❑ *the data collection is sufficient with some areas of reservations;* ❑ *the data collection is sufficient for the requirements;* ❑ *the data collection significantly exceeds requirements; or* ❑ *there is insufficient information to judge the suitability of this characteristic.* *Comments:* _____ _____ _____ _____ _____ _____ _____ _____